12-2005

The Bovine Coronavirus 2'-*O*-Methyltransferase Binds *Cis*-Acting Stem-Loop IV In the 5-Prime Untranslated Region Of The Viral Genome

Tara Beth Tucker
University of Tennessee, Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Tara Beth Tucker entitled "The Bovine Coronavirus 2'-*O*-Methyltransferase Binds *Cis*-Acting Stem-Loop IV In the 5-Prime Untranslated Region Of The Viral Genome." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Microbiology.

David A. Brian, Major Professor

We have read this thesis and recommend its acceptance:

Melissa Kennedy, Tim Sparer

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Tara Beth Tucker entitled "The Bovine Coronavirus 2'-*O*-Methyltransferase Binds *Cis*-Acting Stem-Loop IV In The 5-Prime Untranslated Region Of The Viral Genome." I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Microbiology.

David A. Brian

David A. Brian, Major Professor

We have read this thesis
and recommend its acceptance:

Jeri E.

Melissa Kennedy

Accepted for the Council:

Vice Chancellor and Dean of
Graduate Studies

THE BOVINE CORONAVIRUS 2'-*O*-METHYLTRANSFERASE BINDS *CIS*-ACTING STEM-LOOP IV IN THE 5-PRIME UNTRANSLATED REGION OF THE VIRAL GENOME

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Tara Beth Tucker

December 2005

ACKNOWLEDGEMENTS

There are a number of people to whom I would like to extend my thanks and appreciation. Many thanks go to my mentor, Dr. David Brian, for his guidance, support, and encouragement, and for being such a wonderful and enthusiastic teacher. I would like to thank Dr. Hung-Yi Wu, Dr. Kimberley Nixon, Dr. Agnieszka Dziduszko, and Kortney Gustin for many valuable discussions and helpful advice. I would also like to thank my other lab mates Bo-Jhih Guan, and Gwyn Williams for their help and friendship.

I want to extend appreciation to my committee, Dr. Melissa Kennedy and Dr. Tim Sparer for their time, effort, and concern.

Lastly, I want to recognize and thank my parents Donald, Sr. and Mary Tucker for their unending support, and my brother Dr. Donald Tucker, Jr. for being my longstanding role model, and for leading me to science and medicine.

ABSTRACT

The positive-stranded coronavirus genome, at 32 kilobases in length, is the largest known viral RNA genome, and internal *cis*-signaling elements directing its replication have been described only within the last ten years. The bovine coronavirus genome encodes 26 proteins in the region between the 5'-terminal 210-nt untranslated region and the 3'-terminal 298-nt untranslated region. Here, genes for 5 of the 26 proteins were cloned into bacterial expression plasmids for the long-term goals of characterizing enzymatic and RNA binding properties. These genes encode enzymes postulated to interact directly with the *cis*-acting RNA elements and carry out RNA synthesis, namely, the RNA-dependent RNA polymerase, the helicase, the exonuclease, the endonuclease, and the 2'-*O*-methyltransferase. For a detailed analysis, bacterially-expressed BCoV 2'-*O*-Methyltransferase was purified and (i) tested for enzymatic activity, which is presumably a 2'-*O*-methylation of 5'-terminal cap structures, and (ii) tested for its binding to terminal genomic regions known to contain *cis*-acting replication elements. Methyltransferase activity was not found, suggesting the proper conditions were not met or the proper template was not used, or perhaps, as with many viral enzymes made from a polyprotein precursor, it does not function as a unit-length molecule. Using the electrophoretic mobility shift assay, the 2'-*O*-Methyltransferase was found to bind *cis*-acting stem-loop IV in the 5' untranslated region, but does not bind other *cis*-acting elements, including the region in gene 1 containing stem-loops V and VI or the 3'-proximal *cis*-acting bulged stem-loop and pseudoknot. The results of this study suggest that the putative bovine coronavirus 2'-*O*-Methyltransferase uses stem-loop IV as a binding site to carry out methyltransferase function(s) yet to be discovered.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2'-*O*-MT	2'-*O*-Methyltransferase
BCoV	bovine coronavirus
DI RNA	defective interfering RNA
EMSA	electrophoretic mobility shift assay
EndoN	endonuclease
ExoN	exonuclease
Hel	helicase
HRT cells	human rectal tumor cells
kb	kilobase
kDa	kilodalton
MHV	mouse hepatitis coronavirus
MT	methyltransferase
N protein	nucleocapsid protein
nt	nucleotide
ORF	open reading frame
RdRP	RNA-dependent RNA polymerase
SARS-CoV	severe acute respiratory syndrome coronavirus
sg mRNA	subgenomic messenger RNA
UTR	untranslated region
wt	wild-type
nsp	nonstructural protein

CHAPTER I

LITERATURE REVIEW ON REPLICATION PROTEINS OF THE BOVINE CORONAVIRUS

The Coronavirus Family

The family *Coronaviridae* (9) is one of 17 families of animal RNA viruses, each distinguished by one or more unique feature in structure or replication strategy. The *Coronaviridae* belong to the order Nidovirales, to which two other families, the *Arteriviridae* and the *Roniviridae,* also belong (10). Coronaviruses are enveloped, single-stranded, positive-sense RNA viruses possessing a genome of 28-32 kilobases in length, the largest RNA genome known for any RNA virus. Coronaviruses are divided into three groups named Groups 1, 2 and 3, based on antigenic differences. Genome sequence differences support this classification. There is continuing debate with regard to which group the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) belongs. Recent evidence suggests it most closely fits with Group 2 coronaviruses of which the bovine coronavirus (BCoV) is a member (Figure I-1) (32). Long before the discovery of the SARS-CoV in the spring of 2003, much research had been done on the replication strategy of coronaviruses and on the pathogenesis of many important human and animal coronaviruses. These included the avian infectious bronchitis virus, the porcine transmissible gastroenteritis virus, the bovine coronavirus, the feline infectious peritonitis virus, the mouse hepatitis viruses and the human respiratory coronaviruses. The mouse hepatitis coronavirus has been especially intensely studied as a model for coronaviruses in animals and it was with this virus that many of the intriguing features of coronaviruses were first described. These include extremely high rates of recombination among viral

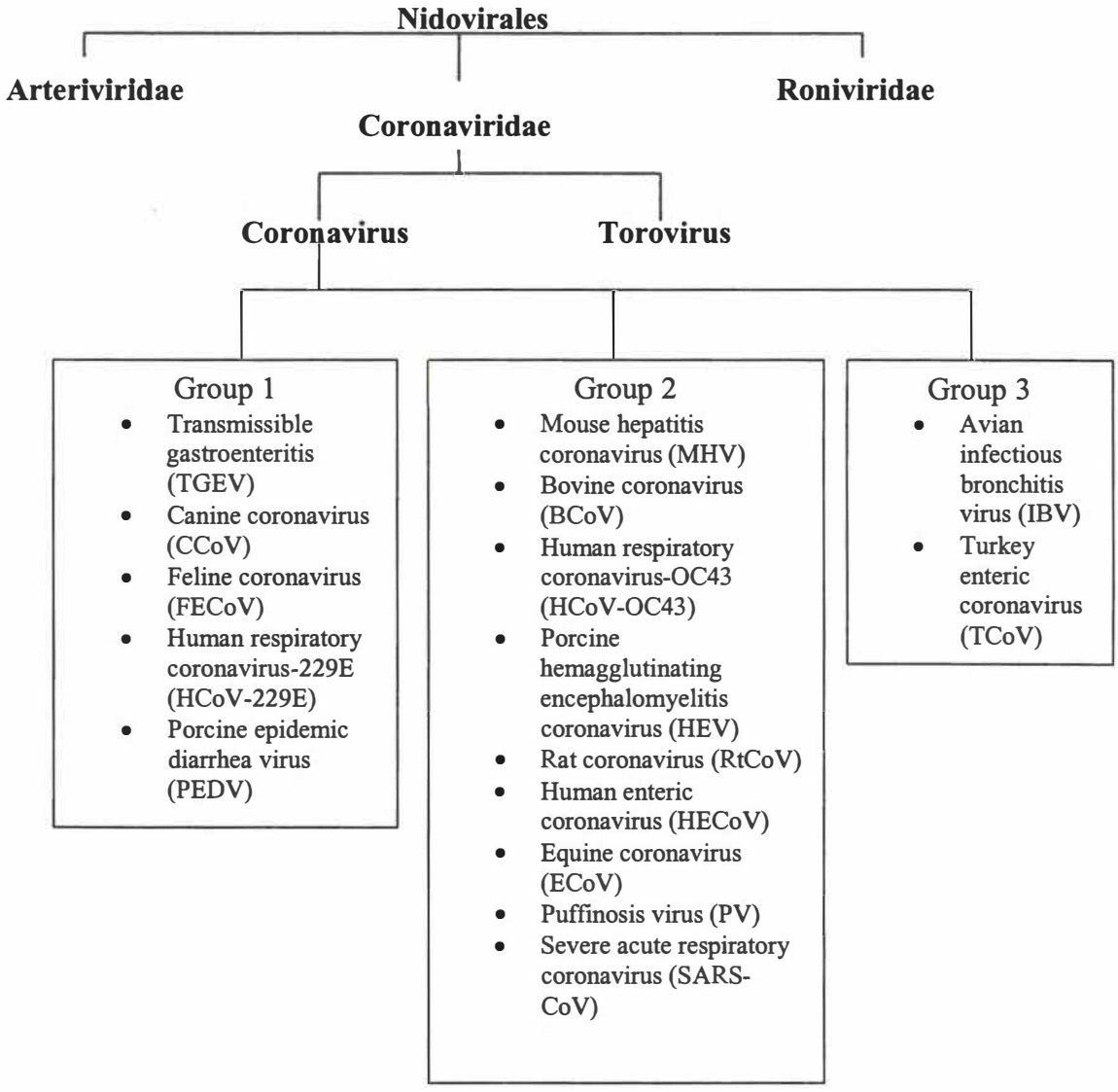

Figure I-1. Taxonomy of coronaviruses. Coronaviruses are divided into three groups as depicted here. Representative members of each group are also shown. The placement of the SARS coronavirus is still under debate, with some investigators proposing it belongs to an entirely new group, Group 4 (32).

strains (approaching 10%) and long-term viral persistence in the host (in some cases the life of the mouse, over 700 days) that often accompanies a demyelinating encephalitis (18) (hence making this an experimental model for multiple sclerosis) (34). With the feline infectious peritonitis virus, immune-mediated disease was found (13). Since the discovery of the SARS-CoV, research on coronaviruses has intensified in part because of its severity as a human pathogen and the consequential need for new vaccines and therapeutic agents. Intensified research has also led to the discovery of SARS-CoV-like viruses in horseshoe bats in Southeast Asia. The identification of key steps in the coronavirus life cycle is now more important than ever as they represent potential target sites for drug design. Among the challenges is a detailed characterization of the enzymes involved in coronavirus genome replication.

Coronavirus Replication Strategy

The genome of the coronavirus, as with most positive-strand RNA viruses, upon entry into a cell acts as mRNA for synthesis of the proteins responsible for virus replication. In the coronavirus, the proteins with enzymatic function responsible for genome replication appear to all come from open reading frame (ORF) 1, also called the replicase gene (Fig. I-2). (Although this is technically a polycistronic gene and the individual products of the gene are derived by proteolytic processing of the ORF 1 gene product, for ease of reference in this thesis, ORF 1 will be considered to be comprised of 16 separate genes, each encoding a separate protein.) ORF 1ab is translated through a -1 ribosomal frameshifting event which occurs at the slippery sequence, UUUAAAC, present at the junction of ORFs 1a and 1b. From ORF 1a comes two proteases, the minor

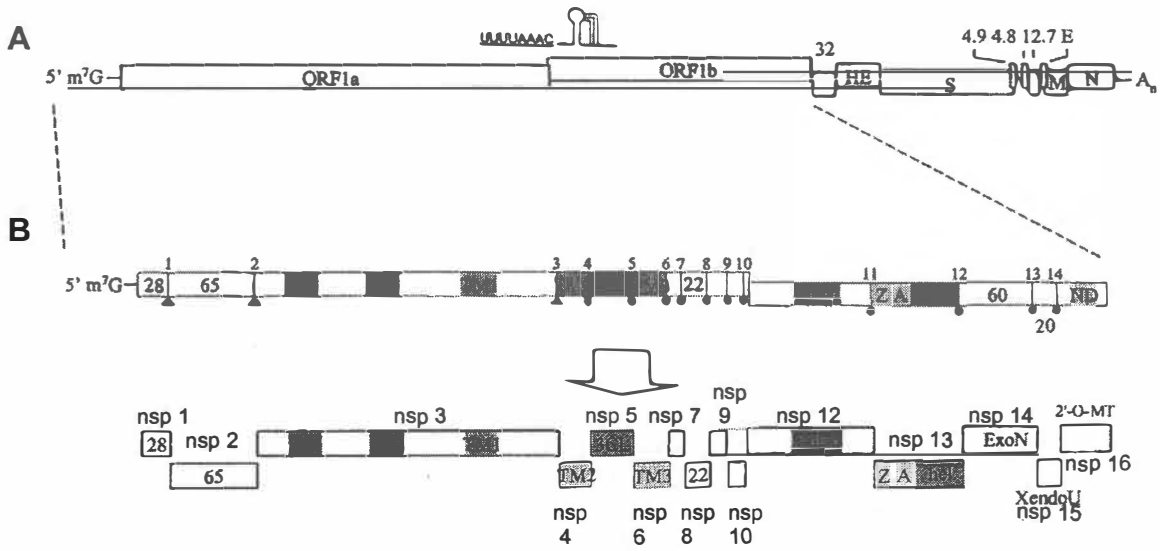

Figure I-2. Schematic representation of the BCoV genome. *A*. The full-length BCoV genome. This depicts ORF 1 (the fused ORF 1a and ORF 1b) also called the replicase gene in the 5' proximal two-thirds of the genome, and the structural protein genes in the 3' proximal one-third of the genome. *B*. A close-up representation of the replicase gene showing the positions of regions encoding the nonstructural proteins. These are made as a polyprotein that are proteolytically processed by viral proteases encoded within ORF 1a. (For ease of description in this thesis, the short regions encoding nsp 1-16 are called genes.) The numbers inside the boxes representing nsp 1, nsp 2, and nsp 8 refer to the molecular weight of the protein products. A very small protein of 14 amino acids exists at the 3' end of ORF 1a and is called nsp 11(7, 32).

protease, a papain-like protease (PLP-1 and PLP-2), and the main protease, a picornavirus 3C-like serine protease (3CL), that together cleave the ORF 1a and 1ab polyproteins into the 16 nonstructural protein products (named nsp 1 through nsp 16). From ORF 1b come five proteins (nsp 12 through nsp 16), all enzymes, predicted to play a direct role in RNA replication and transcription (i.e., the synthesis of sg mRNAs). These are identified by viral genomics analyses to be the RNA-dependent RNA polymerase (RdRP) (nsp 12), the helicase (Hel) (nsp 13), the 3' to 5' exonuclease (ExoN) (nsp 14), the poly(U)-specific endoribonuclease (EndoN) (nsp 15), and the *S*-adenosylmethionine-dependent ribose 2'-*O*-Methyltransferase (2'-*O*-MT) (nsp 16) (Figure I-2) (32). These proteins are presumably used to replicate the viral genome (replication) and to make a 3'-coterminal nested set of subgenomic mRNAs (transcription), processes which occur in the cytoplasm in as yet uncharacterized membrane-associated replication and transcription complexes. In coronaviruses (as in arteriviruses, the only other family known to do it), the transcription process is uniquely discontinuous since a 5'-leader sequence (ranging from 65 to 93 nts in length in coronaviruses), encoded only at the 5'-terminus of the genome, gets placed onto the 5'-end of each subgenomic mRNA. From these subgenomic mRNAs, only the 5'-most open reading frame (ORF) is translated to give rise to viral structural proteins. The coronavirus structural proteins include the spike glycoprotein (S), the envelope protein (E), the membrane glycoprotein (M), and the nucleocapsid phosphoprotein (N). Various other proteins (2-5) of unknown function, often referred to as luxury or nonstructural proteins, are also made (4, 39).

The Bovine Coronavirus (BCoV) Model System

The molecular biology of bovine coronavirus replication was initially studied in our laboratory because BCoV was one of the few culturable group 2 coronaviruses at the time (1976) that caused natural disease in the gastrointestinal tract. The aim at that time, as now, was to learn the details of virus replication in the hope of developing anti-coronaviral chemotherapeutic agents. The molecular biology of BCoV was therefore studied in parallel with a few other key coronaviruses, primarily the mouse hepatitis coronavirus (MHV) and porcine transmissible gastroenteritis virus, and fundamental discoveries about coronavirus replication were made in all three viruses (9). The unique contributions made with BCoV were (i) a molecular description and characterization of the Hemagglutinin-Esterase structural glycoprotein, a protein phylogenetically related to the HE protein in group C influenza viruses, that appears to contribute to the neurotropism of some group 2 coronaviruses and that is not made in most strains of MHV or in any virus in groups 1 and 3 (21, 24). (ii) confirmation of the presence of subgenomic mRNA minus strands (that contributed in a major way to the current model of coronavirus transcription) (20). (iii) the discovery and characterization of a simple defective interfering RNA (the simplest of all coronavirus DI RNAs) that is comprised of the two ends of the viral genome (the 5' 498 nts and the 3' 1635 nts) and that replicates in the presence of the parent helper virus (5). This 2.2 kilobase DI RNA has been a major tool in the discovery and characterization of the *cis*-acting elements of coronavirus genome replication (6, 30, 31, 36, 37).

Cis-Acting RNA Elements for Genome Replication in BCoV

Seven *cis*-acting RNA replication elements have been described to date for the 2.2 kilobase BCoV DI RNA, a minigenome, that presumably are also required for the replication of the full-length viral genome (Figure I-3). (i) The 5'-terminal 84 nts in the BCoV genome which contains two stem-loops, named stem-loops I and II, is a sequence required for minigenome replication (6). However, since these stem-loops are not highly conserved in coronaviruses, even among the group 2 coronaviruses, it is not likely that the higher-order structures, per se, are required for the *cis*-acting function. (ii) The 5'-proximal stem-loop III is a highly-conserved higher-order structure in all coronaviruses, based on *mfold* predictions by the Zuker algorithm (40), that is required as a higher-order structure in both the positive and negative strands for minigenome replication (30). (iii) The 5'-proximal stem-loop IV is a highly-conserved higher-order structure in group 2 coronaviruses that has a homolog in all coronaviruses and is a required higher-order structure in the positive strand for minigenome replication (31). (iv) The 5'-proximal stem-loop VI is a higher-order structure mapping within the nsp 1 coding region that is conserved in group 2 coronaviruses and is required as a higher-order structure in minigenome replication (Brown, Nixon, Senanayake and Brian, manuscript in preparation). (v) The 3'-proximal bulged stem-loop and adjacent pseudoknot that map just downstream of the N stop codon within the 3' untranslated region function together as a *cis*-acting element in minigenome replication (36). This element is highly-conserved in group 2 coronaviruses and has been shown to function as a *cis*-acting element in the MHV minigenome and genome as well (17). (vi) The 3'-proximal octameric sequence GGAAGAGC found in all coronaviruses sequenced to date is part of a predicted stem-

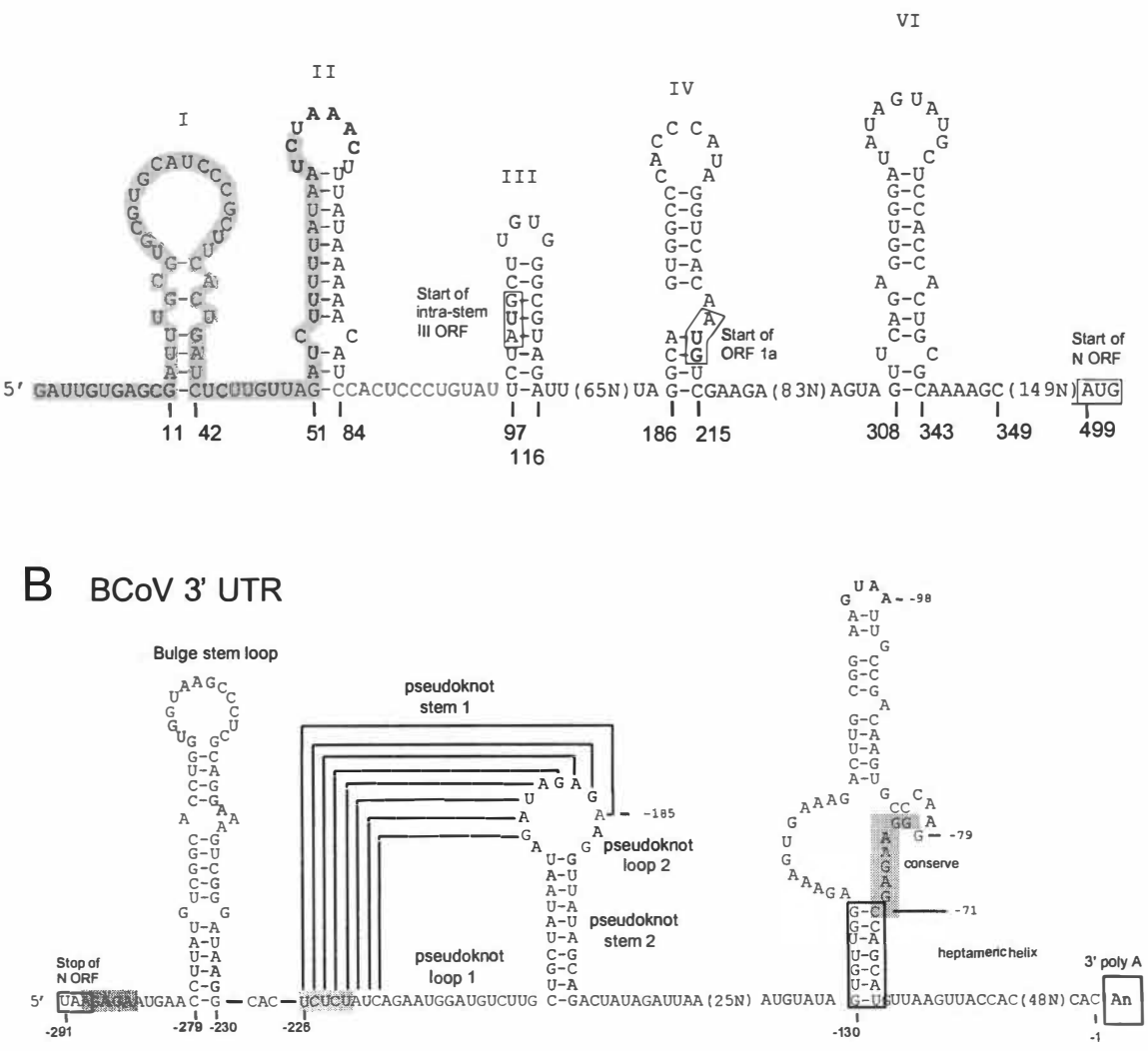

Figure I-3. Schematic representation of the seven *cis*-acting replication elements in the 5' and 3' termini of the BCoV genome and in the nsp 1 coding region. *A*. BCoV 5' UTR and part of the nsp 1 coding region depicting (i) stem-loops I and II (which together function as a *cis*-acting element), (ii) stem-loop III, (iii) stem-loop IV, and (iv) stem-loop VI. *B*. BCoV 3' UTR depicting (v) 5'-proximal bulged stem loop and adjacent pseudoknot (which together function as a *cis*-acting element), (vi) 3'-proximal bulged stem-loop with the internal GGAAGAGC octameric sequence, and (vi) poly (A) tail. Numbers represent nucleotide positions from the respective termini.

loop in coronaviruses in groups 1, 2 and 3 and is required as a *cis*-acting sequence in the replication of the BCoV minigenome (Wu and Brian, unpublished). (vii) The 3'-terminal poly(A) tail is a *cis*-acting element in minigenome replication (33).

It is presumed that the *cis*-acting RNA elements function through interactions with viral and, or, cellular proteins. To date, however, only a few of these interactions have been identified, and, where identified, it is not yet been established whether the RNA-protein interactions are required for the *cis*-acting function. In relation to the seven identified *cis*-acting elements in the BCoV (and closely-related MHV) system, the following can be said. (i) It has been shown in MHV that the N protein binds the UCUAAAC leader associated transcription-regulating core sequence within the first 84-nt *cis*-acting element with high affinity (K_D=14μM) (27) and this will probably hold true as well for BCoV. (ii) Regarding stem-loop III in BCoV, it has been shown that the viral N protein and unidentified viral proteins of 22 and 38 kDa bind the stem-loop in the positive strand and that seven cellular proteins in the molecular weight range of 76 and 25 kDa of unknown identity bind stem-loop III in the negative strand (Raman and Brian, submitted). (iii) Regarding stem-loop IV in BCoV, it has been shown that eight cellular proteins in the molecular weight range of 78 to 25 kDa of unknown identity bind the stem-loop in the positive strand in a higher-order-dependent manner (31). (iv) Regarding stem-loop VI in the nsp 1 coding region of BCoV, it has been shown that p28, the protein product of the nsp 1 coding region, binds to some region within this coding region but it is not known whether it binds to stem-loop VI per se (Kortney Gustin, unpublished data). (v) Regarding the bulged stem-loop and the pseudoknot in the 3' untranslated region of MHV and BCoV, no protein has yet been shown to bind this

region (Agnieszka Dziduszko, unpublished data). (vi) Regarding the GGAAGAGC octameric region and the associated stem-loop, no protein has yet been shown to bind this region except for the exciting new observation described below that the bovine methyltransferase binds at or near the GGAAGAGC octamer (Agnieszka Dziduszko and Tara Tucker, unpublished data). (vii) Regarding the 3'-terminal *cis*-acting poly(A) tail in MHV and BCoV, it has been shown that the poly(A)-binding protein binds this region (33).

It is our hypothesis that the *cis*-acting elements for bovine coronavirus genome replication are binding targets for the enzymes used in genome replication. In that light, I have, along with testing the enzyme activity of the BCoV 2'-O-MT (described below), investigated its potential binding to the 5' untranslated region as a whole, and stem-loop IV in particular. In collaboration with Kortney Gustin, its potential binding to the nsp 1 ORF and in collaboration with Agnieszka Dziduszko, its potential binding to the 3'-proximal bulged stem-loop and adjacent pseudoknot, and the octamer-containing stem-loop, were also tested.

BCoV Replication Enzymes

Putative functions have been assigned to the cleavage products of ORF 1b based on sequence homologies and motif similarities of known RdRP, Hel, ExoN, EndoN, and MT enzymes (7, 32). Since these proteins are only translated when the ribosomal frameshifting event occurs (about 20%-30% of the time *in vitro*), it is suggested that they are needed in much less abundance than the other proteins encoded by the genome (39). The replicase proteins are thought to assemble into a replication complex that is

associated with cellular membranes. Through the action of this complex, the viral genome is replicated and the 3'-coterminal nested set of subgenomic mRNAs is made. While it is presumed that the RdRP polymerizes nucleotides into RNA, the exact functions of the Hel, ExoN, EndoN, and MT are not known. It is speculated that the ExoN may be involved in RNA proofreading and repair functions but these would be novel properties for an RNA virus; however, because of their large genomes, it's possible they may have acquired this ability (39). The MT may be involved in 5' capping, a process thought to increase mRNA stability and aid in ribosomal binding. The BCoV helicase protein contains a putative N-terminal zinc-binding domain, a domain required for helicase function. RNA helicases unwind double-stranded RNA through the hydrolysis of nucleoside triphosphates (39). The helicase may also participate in the 5' capping reaction of viral mRNAs since it possesses a predicted 5'-triphosphatase activity (23).

Rationale for a Focus on the Enzymatic and RNA Binding Properties of BCoV 2'-*O*-Methyltransferase

The 2'-*O*-methyltransferase has been implicated in the capping reaction of the 5' end of viral mRNAs, as was shown for an analogous flavivirus enzyme, the Dengue virus NS5 protein (8). A conserved motif of amino acid residues present in the Dengue virus 2'-*O*-MT ($NS5MTase_{DV}$), K61-D146-K181-E217, is also present in the BCoV 2'-*O*-MT, K46-D130-K170-E203. Additionally, an analogous motif is also present in another 2'-*O*-MT, VP39 in the double-stranded DNA vaccinia virus, and is identified as K41-D138-

K175-E207. These similarities led us to investigate the properties of the BCoV enzyme further.

The binding pattern of the MT is of interest since it may suggest its site of action and perhaps also other function(s). To study the binding capacity of the BCoV MT, the electrophoretic mobility shift assay (EMSA) was used. By using RNA transcripts containing the individual *cis*-acting elements from the 5' and 3' termini as radiolabeled probes, the binding sites for the MT were sought.

The enzymatic function and binding capacity of BCoV MT have not yet been demonstrated. We were therefore led to address the following questions. With regard to enzymatic activity, will the MT transfer a methyl group from its donor, *S* –adenosyl-L-methionine (SAM), to a 5'-terminal RNA transcript produced *in vitro*? Will the MT bind a potential RNA target with enough affinity and specificity to produce a gel shift in an EMSA? What *cis*-acting replication elements are required for the binding of the MT to BCoV RNA?

CHAPTER II

CLONING OF THE BOVINE CORONAVIRUS RNA-DEPENDENT RNA POLYMERASE, HELICASE, EXONUCLEASE, ENDONUCLEASE AND 2'-*O*-METHYLTRANSFERASE GENES, AND EXPRESSION AND PURIFICATION OF THE 2'-*O*-METHYLTRANSFERASE

Introduction

There are five genes in the BCoV genome that, on the basis of bioinformatic predictions, encode enzymes for RNA metabolism and they are all located within ORF 1b (32). These are, in order, the RNA-dependent RNA polymerase (RdRP), the Helicase (Hel), the Exonuclease (ExoN), the Endonuclease (EndoN), and the 2'-*O*-Methyltransferase (MT). A required -1 ribosomal frameshift at the beginning of ORF 1b enables the synthesis of the ORF 1b proteins, all made in less abundance than the proteins of ORF 1a and in far less abundance than the (mostly) structural proteins made from downstream ORFs by individually-produced subgenomic messenger RNAs.

The RdRP functions to replicate the genome and to synthesize subgenomic mRNAs. The functions of the helicase, exonuclease, and endonuclease are not yet known; however, the helicase may be responsible for unwinding secondary structure of the RNA in order for it to be more accessible to the RdRP. The MT, based on analogy, probably plays a role in 5' capping of the genome and subgenomic mRNAs, which helps stabilize the mRNAs and aids in binding of the mRNAs to the ribosome. As a 2'-*O*-methyltransferase, its probable role is in catalyzing the transfer of a methyl group from SAM to the first nucleotide 3' of the triphosphate bridge of the cap structure (8).

Each of the five BCoV replicase genes were cloned into the TOPO-XL vector (Invitrogen) and subsequently into pGex (Amersham) and pET-28a (Novagen)

expression vectors to optimize chances for protein production. Additionally the MT was cloned into pET-SUMO (Invitrogen). The long-term objectives were to express and purify each protein for enzyme function analyses, RNA binding analyses and ultimately crystallization and structural analyses (the last goal is in collaboration with the laboratory of Dr. Christopher Dealwis, Department of Biochemistry, University of Tennessee, Knoxville). In addition, plasmids were sent to Proteintech Group, Inc. (Chicago, IL) for the commercial production of rabbit polyclonal antiserum to each. Three separate expression vectors were used in this study in order to determine the optimal conditions for protein expression. The pGex vector expresses a fusion protein with glutathione-*S*-transferase (GST), the pET-28a vector expresses a His-tagged fusion protein, and the pET-SUMO vector yields a non-fused protein after protease cleavage from the SUMO protein.

Materials and Methods

Virus and cells

Human rectal tumor (HRT) cells were infected with the bovine coronavirus-Mebus strain at an MOI of 1. HRT cells were grown as monolayers in Dulbecco's modified Eagle medium containing 10% fetal bovine serum (Hyclone). At 6 hours post infection, total RNA was harvested using TriZol (Invitrogen) and purified by phenol-chloroform extraction. cDNA for the RdRP was made by using reverse transcriptase and the primer 3.1-268RL. For the Hel gene, primer 18302(+) was used, for the ExoN gene, primer 19973(+) was used, for the EndoN gene, primer 21095(+) was used, and for the MT gene, primer 21992(+) was used (Table II-1).

Table II-1. Oligonucleotides used in this study.

Oligonucleotide[a]	Sequence[b]	Binding Region (nt position)[c]
3.1-268RL(+)	GCCAGTTGCCTTATATTTG	17996-18014
18302(+)	GCTGTAACCATCTCTATCAGCAAAC	18278-18302
19973(+)	TGTGTCTCCAACCTTGTCCA CCACTACGCC	19934-19973
21095(+)	AACACTGCCACCCAGAGCTAAC	21074-21095
21992(+)	TGAGGGGTTGATAGTGATTTTATAATTG	21965-21992
BamRdRP(-)	CGGGATCCTCAAAAGATACTAA TTTTTTTAAACGGGTTCGGGG	13318-13351
Pst1RdRPStop(+)	CGGCTGCAGTCATCATCACTGCATAAC TGCACTTCTTAAATACATGTTCTTG	16067-16100
BamHel5(-)	CGGGATCCAGTGTTGGAGC TTGCGTGGTCTGCTC	16101-16126
EcoRIHelStop(+)	CGGAATTCTCATCATCATTGAACT CTCGTTTCAACGGCCTGTGGC	17882-17909
BamExo5(-)	CGGGATCCTGTAGTACCAATTTATTTA AAGATTGTAGCAAGAG	17910-17934
EcoRIExoStop(+)	CGGAATTCTCATCATCATTGTAGC TTGGTGAACGTATTCCCAC	19472-19449
BamEndo5(-)	CGGGATCCAGCTTGGAGA ATGTTGTATATAA	19473-19495
EcoRIEndoStop(+)	CGGAATTCTCATCATCATTGCAAA CGAGGATAGAAAGTC	20574-20594
BamMT5(-)	CGGGATCCGCTGCATCTGACTGGA AGCCTGG	20595-20617
MfeMTStop(+)	CGCAATTGTCATCAGATTACATTAAC CATACTGTCACCAAC	21465-21491

[a]Oligonucleotide binds to either plus-sense RNA as indicated by (+), or to minus-sense RNA as indicated by (-).
[b]Underlined bases indicate differences from genomic sequence.
[c]Numbers correspond to Bovine Coronavirus-Mebus strain genome sequence.

Cloning into TOPO-XL plasmid vector

The TOPO-XL vector (Invitrogen) was designed to facilitate the cloning of PCR products since it takes advantage of the extra terminal 3' deoxyadenosine residue added by the Taq polymerase during PCR. For cloning, cDNAs were used as templates in PCR reactions with Taqmaster polymerase (Eppendorf) along with the appropriate primers to amplify each gene (Table II-1). The primers used were engineered to contain restriction enzyme sites at the 5' and 3' ends to aid in cloning, and stop codons at the 3' ends to aid in protein expression. For ExoN and EndoN clones, a *BamH* I site was engineered at the 5' end and an *EcoR* I site at the 3' end. Additionally, three stop codons were engineered into each 3' primer. This was necessary since the genes are expressed naturally as internal segments of a polyprotein subsequently cleaved into individual functional proteins by ORF 1a-encoded proteases. Primers used to amplify the ExoN gene were BamExo5 and EcoRIExoStop. Primers used to amplify the EndoN gene were BamEndo5 and EcoRIEndoStop. To clone the 2'-*O*-MT, the 5' primer contained a *BamH* I site and the 3' primer a stop codon and an *Mfe* I site, as the sequence of the gene already contained an *EcoR* I site. The primers used were BamMT5 and MfeMTStop. Digestion with *Mfe* I and *EcoR* I produce compatable sticky ends so the *Mfe* I-digested 2'-*O*-MT gene ligates readily into an *EcoR* I-digested vector. To clone the RdRP, PCR was performed using primers BamRdRP(-) and PstIRdRPStop, and for the Hel, primers BamHel5 (-) and EcoRIHelStop were used. An extra T was designed into the -1 frameshift of the RdRP 5' primer to ensure a correct reading frame for the gene. Each PCR reaction product was purified by electrophoresis in a crystal violet-containing agarose gel (Invitrogen), and each product was ligated into the TOPO-XL vector using

the recommended procedure of the supplier (Invitrogen). The ligation mixtures were transformed into TOP10 electrocompetent cells and cells were plated on 2XYT agar plates in the presence of kanamycin and grown overnight at 37°C. Colonies were screened by PCR using primers that bound within the cloned gene and positive colonies were used to make plasmid minipreparations with the Wizard Miniprep system (Promega). Sequences of cloned genes were determined by automated DNA sequencing in the Molecular Biology Resource Facility at the University of Tennessee.

Cloning into expression vectors

Three of the replicase enzymes, ExoN, EndoN, and MT, were cloned into expression vectors. The RdRp and Helicase genes in the TOPO-XL vector contained unwanted point mutations that were not easily repaired by repeated cloning so these were not used futher. For the ExoN and EndoN, 10µl of each respective TOPO-XL plasmid minipreparation was double-digested with *BamH* I and *EcoR* I at 37°C for two hours. The digested products were then resolved by electrophoresis on a 1% agarose ethidium bromide gel and products isolated with the Geneclean III kit (Q·Biogene). Expression vectors, pGex and pET-28a, were also double-digested with *BamH* I and *EcoR* I and the products dephosphorylated with calf intestinal phosphatase (New England Biolabs), resolved by electrophoresis in a 1% agarose ethidium bromide gel, and isolated with the Geneclean III kit (Q·Biogene). The purified linearized dephosphorylated vectors and purified digested gene products were ligated as required using the Quick Ligase kit (New England Biolabs). Ligated products were transformed into chemically competent DH5α cells (Invitrogen) and 100µl of each transformation reaction mixture was plated onto

2XYT agar plates containing ampicillin in the case of the pGex constructs and containing kanamycin in the case of the pET-28a constructs, and plates were incubated overnight at 37°C. Colonies were screened by PCR and positive colonies were used to make plasmid minipreparations which were then sequenced to verify a wild-type sequence in the clone.

The MT gene was similarly cloned into pGex and pET-28a vectors except that *Mfe* I replaced *EcoR* I for digestion of the MT-TOPO-XL plasmid minipreparation. Vectors were linearized with *BamH* I and *EcoR* I and ligation, transformation, plating, screening, and sequencing procedures were as described above.

Additionally, the MT gene was cloned into pET-SUMO to produce a product with no extraneous fusion sequence following the SUMO protease cleavage step. For this, a 5' primer that corresponds to the precise 5' end of the MT gene (i.e., without a restriction endonuclease site) was used. The 3' primer contained two sequential stop codons. The PCR product was isolated using Q·Biogene's Geneclean III kit and ligated into the pET-SUMO vector following the protocol of the manufacturer (Invitrogen). The ligation mixture was transformed into One Shot Mach1-T1 chemically competent *E. coli* cells (Invitrogen) and 100µl of the transformation reaction mixture per plate was spread on 2XYT agar plates containing kanamycin. Screening and sequencing procedures were carried out as described above.

Bacterial expression and purification of cloned BCoV 2'-*O*-Methyltransferase

MT was expressed from pGex-MT as a GST fusion protein, and from pET-28a-MT as a his-tagged MT in order to increase the chances of getting efficient protein expression. For expression from pGex-MT, plasmid DNA was transformed into BL21

pLysS *E. coli* cells and grown in TB medium with ampicillin and chloramphenicol. Expression was induced with IPTG at 15˚C overnight. Cells were lysed and proteins in the lysate bound to glutathione-sepharose resin (Amersham Biosciences) by incubation for two hours and end-over-end rotation. The column was washed sequentially with thrombin wash buffer (50mM Tris pH 8.0, 200mM NaCl, 5% glycerol), high salt buffer (50 mM Tris pH 8.0, 0.5 M NaCl), CHAPS buffer (50mM Tris pH 8.0, 200mM NaCl, 8mM CHAPS), and then standard wash buffer to remove salt and CHAPS. Thrombin was added to the column and the mixture incubated at room temperature for 2 hours with occasional rotation. The flowthrough was collected and the resin washed with thrombin wash buffer to collect any remaining protein. The eluate was subjected to Superdex75 column chromatography to increase protein purity.

For expression from pET-28a-MT, plasmid DNA was transformed into BL21 pLysS *E. coli* cells and grown in TB with chloramphenicol and kanamycin. Expression was induced with IPTG and grown for 15 hours at 15˚C, and cells were pelleted and flash frozen in liquid nitrogen. Thawed cells were resuspended in lysis buffer (50mM Tris pH 8.0, 100mM NaCl, 1mM EDTA, 10% glycerol, 5mM DTT, 1mM PMSF, 1ul Benzonase per 10g cells) and the lysate was clarified by centrifugation for one hour. Lysate was incubated with NiNTA resin (Qiagen) at 4˚C for one hour on an orbital shaker. The lysate-resin mixture was poured into an EconoColumn and flowthrough collected. The column was washed with ~50 column volumes of wash buffer and thrombin was added at a concentration of ~5μL per liter of cells. The reaction mixture was incubated at room temperature for one hour with gentle agitation every few minutes, after which the

flowthrough plus two column volumes of wash were collected. The MT contained in the flowthrough was purified on a Superdex75 size exclusion column.

Synthetic oligonucleotides used in this study

Oligonucleotides used in this study are described in Table II-1. All oligonucleotides were synthesized by Invitrogen Life Sciences (Carlsbad, CA).

Results

Cloning of the BCoV RNA-dependent RNA polymerase, Helicase, Exonuclease, Endonuclease, and 2'-*O*-Methyltransferase genes into the TOPO-XL vector

The procedures used for cloning the five replicase genes into the TOPO-XL vector are described in Materials and Methods. Figure II-1, *A* illustrates the PCR products of the ExoN, EndoN and MT genes resolved by agarose gel electrophoresis prior to cloning into the TOPO-XL vector. The sequences of the cloned RdRp and Hel genes showed them to have point mutations at various sites throughout. cDNA cloning of each was repeated several times but wt sequences were not found and thus cloning into expression vectors was not pursued at this time. The sequences of the cloned ExoN, EndoN and 2'-*O*-MT genes, however, were wt and were futher cloned into expression plasmid vectors of *E. coli*.

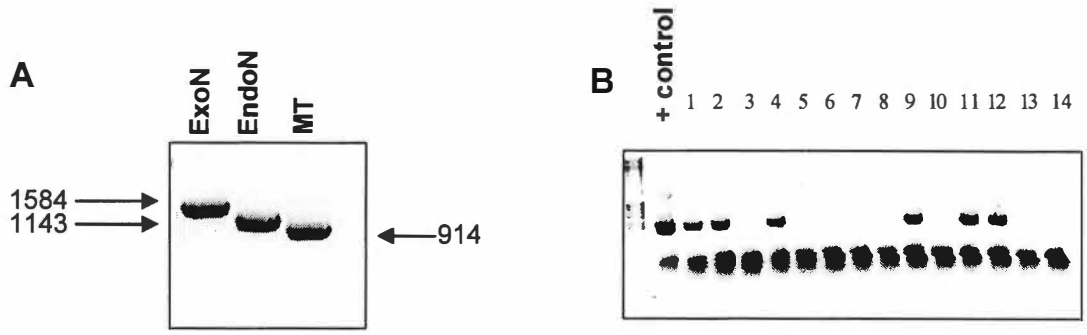

Figure II-1. Cloning the BCoV replication enzymes. *A*. PCR products of the ExoN, EndoN, and MT genes from a cDNA template. The numbers beside the bands indicate the size of the PCR product in kilobase pairs including the stop codons and restriction enzyme sites that were designed into the primers. *B.* PCR screen of the 2'-*O*-MT cloned into the pET-28a vector. Lane 1, + control (1µl of the ligation mixture); Lanes 1, 2, 4, 9, 11, and 12, colonies positive for the 2'-*O*-MT gene; Lanes 3, 5-8, 10, 13 and 14, colonies with no insert.

Cloning of the BCoV Exonuclease, Endonuclease, and 2'-*O*-Methyltransferase genes into expression vectors

Many colonies screening positive for the ExoN, EndoN and 2'-*O*-MT genes were found to contain wt gene sequences and some of these were used as sources for cloning into expression vectors. Mini-prepped plasmid DNA preparations were digested with restriction endonucleases corresponding to the engineered sites and inserts, purified by electrophoresis on a 1% agarose ethidium bromide gel, were ligated into appropriately linearized and dephosphorylated expression vectors that were used to transform chemically competent DH5α cells. Resulting colonies were screened by PCR for inserts of the proper size as depicted in Figure II-1, *B*. The MT was cloned into the pET-28a and pGex vectors forming pET-28a-MT and pGex-MT, respectively, and subsequently cloned into pET-SUMO forming pET-SUMO-MT. The ExoN and EndoN were cloned into pET-28a and pGex (*data not shown*).

Expression and purification of the 2'-*O*-Methyltransferase

For the purposes of measuring enzyme activity (Chapter III) and RNA binding (Chapter IV), the MT expressed from pGex-MT was used since the yield from this vector was the best. Figure II-2 illustrates the MT protein as a fusion protein with GST and the intervening amino acids. Attempts to cleave off the MT from the GST protein with 3C-protease (which cleaves immediately adjacent to the MT) to yield a native MT protein were unsuccessful. Therefore, cleavage with thrombin was used although this approach left an additional 11 N-terminal amino acids on the MT which could possibly interfere with either enzyme function or RNA binding. A colony screening to find positively

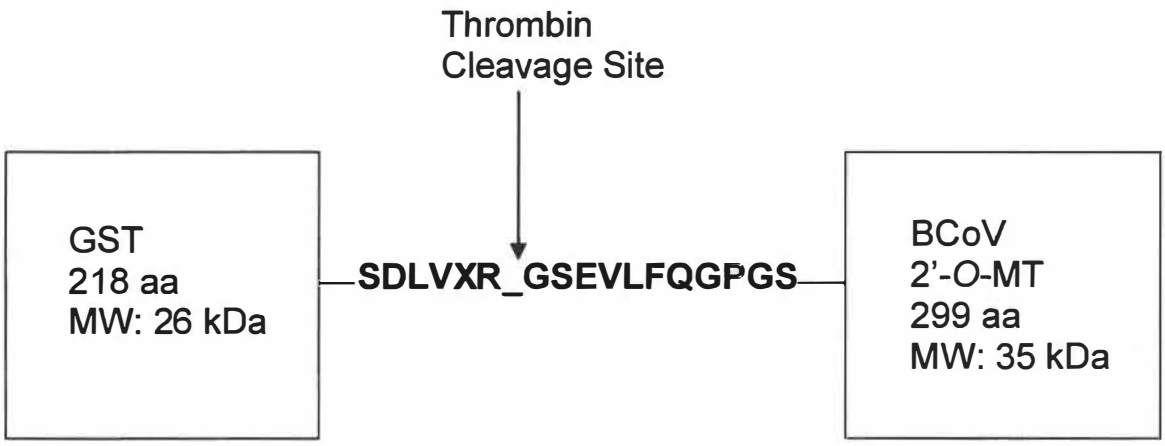

Figure II-2. GST-Fusion construct showing the fusion protein produced when the MT is expressed from the pGex vector (pGex-MT). The arrow indicates the site at which the protease, thrombin, cleaves GST from the target protein. The amino acids to the right of the cleavage site in the diagram remain attached to the released MT.

transformed bacteria is illustrated in Figure II-3, *A*. After confirming expression of an induced fusion protein, cells were lysed and the fusion protein purified by affinity chromatography. MT was cleaved from the fusion protein and collected. Samples of the eluate were resolved by polyacrylamide gel electrophoresis to determine location of the protein (Figure II-3, *B*). A sample from the column resin eluted by SDS-PAGE sample treatment buffer was also resolved by polyacrylamide gel elctrophoresis to determine the degree of MT column retention. The presence of a 35 kDa band in the sample containing the eluate collected after the thrombin cleavage indicated that the MT was expressed in the cells and collected. However, presence of the 35 kDa band in the sample containing the resin indicated that not all of the protein was eluted in the initial step following thrombin cleavage. A ~26 kDa faint band is also present in the sample of the MT eluate indicating that some GST impurities may be present. Expression of the MT from the pET-28a vector was performed (*data not shown*). In this case, however, expression levels were poor and the purity of the protein was low.

Discussion

The principal complication in cloning the five replicase genes from ORF 1b of BCoV was the occurrence of random point mutations in the resulting clones. Random mutations were most prominently found in the constructs of the RdRP and Hel genes. Wild-type clones of the ExoN, EndoN, and MT genes in expression vectors were ultimately obtained. PCR mutagenesis reactions were performed on clones of the RdRP in attempts to repair single point mutations, but in these attempts alternate point mutations were

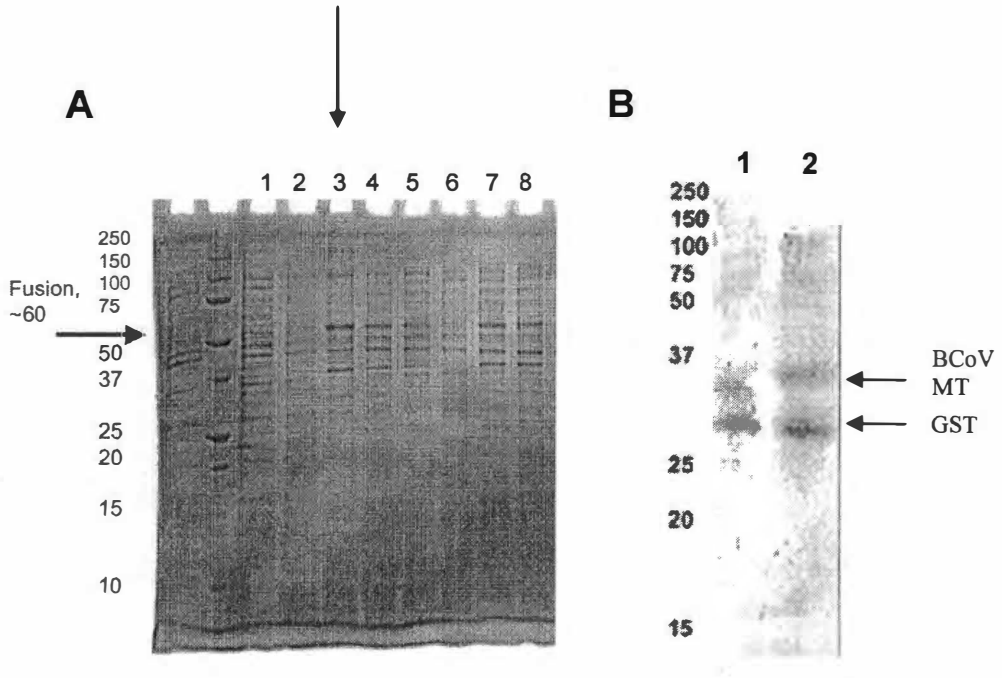

Figure II-3. Expression of the BCoV MT from pGex-MT. *A*. Colony screen of cell colonies transformed with fusion protein-containing plasmid. The fusion protein has a molecular weight of ~60kDa. Lane 3, indicated by the arrow, represents the colony with the best expression of the fusion protein. *B*. SDS-PAGE gel containing a sample of the eluate after the thrombin cleavage (lane 1) and material retained on the resin from the column (lane 2). Arrows indicate the location of the bands representing the MT and GST. The MT is ~35kDa and GST is ~26kDa.

found. Since the point mutations were random and varied in position it is believed they were introduced at a step subsequent to PCR, possibly during selection in the transformed cells. It is postulated that the RdRp and Hel enzymes contain sequences toxic in some fashion to the *E. coli* host and that non-toxic variants are selected by the host. The BCoV 2'-*O*-MT gene was cloned into pGex and used to express the MT as a fusion protein with GST. The 2'-*O*-MT was cleaved from the GST protein and used in subsequent experiments which will be discussed in detail later. There were initial expression problems with the protein. (i) Whereas induction at 37°C for 3 hours showed poor expression, induction at 15°C for 20 hours showed much better expression (*data not shown*). (ii) Whereas much 35kDa protein was retained on the resin after washing with normal elution buffer, post-cleavage elution with CHAPS buffer proved much better (*data not shown*). (iii) The presence of a 26kDa band in the eluate lane (Figure II-3, *B*) indicated that some GST, which should remain on the column, had eluted with the MT. (iv) Initially we expected to use the 3C protease site just N-terminal to the start of the MT protein to cleave MT from the GST protein. However, cleavage efficiency of this construct on the column was extremely low. It was then postulated that perhaps the GST and 2'-*O*-MT domains of the fusion protein were interacting a way that blocks access of the protease to its cleavage site, and thrombin was chosen to make the cleavage. Although this enzyme leaves extra amino acids at the N-terminus of MT, once optimized the protocol yielded milligram quantities of the protein used for further experimentation as described in Chapters III and IV.

CHAPTER III

ASSAY FOR THE ENZYMATIC ACTIVITY OF BACTERIALLY-EXPRESSED BOVINE CORONAVIRUS 2'-*O*-METHYLTRANSFERASE

Introduction

The 2'-*O*-Methyltransferase gene in the coronavirus genome has only recently been identified, and this was based on bioinformatic analyses (32). No functional or structural studies on a coronavirus MT have been published to date and it is toward this end that we have initiated studies on the bovine coronavirus MT.

Many viruses that replicate in the cytoplasm encode their own MT, presumably to enable them to methylate their own translation-enhancing 5'-terminal cap structures that would be otherwise out of reach of the cellular methyltransferases that function in the nucleus. The viral MTs also may have evolved other viral-specific functions required for virus replication. It is quite possible that a viral MT would methylate internal bases on the genome that would serve as some kind of heretofore uncharacterized signal for RNA function, or it could possibly methylate an amino acid to carry out another kind of signaling event. Viral MTs are thought to be attractive targets for chemotherapy since they have structural and biochemical features differing from cellular MTs. For this reason a detailed study of the BCoV MT activity to determine its function was undertaken. A description of our analyses follows.

Materials and Methods

In vitro synthesis of three potential target RNAs for the *in vitro* methylation assay

For assays of the BCoV MT activity, it was presumed that the most likely targets for methylation would be an unmethylated G(5')ppp(5')G cap structure at the natural 5' end of the genome (which by virtue of the common leader strategy of coronavirus transcription is also the 5' terminus of the subgenomic mRNAs), and a partially methylated cap structure, m^7G(5')ppp(5')G. However, a terminus with no cap at all was also possible, so a 5'-terminal target with no cap was also made and tested. To obtain these, a cloned BCoV defective interfering (DI) RNA of the BCoV under control of the T7 RNA polymerase promoter was used in an *in vitro* transcription reaction to make the capped and uncapped target RNA. Construction of the DI RNA-containing plasmid, pDrep1, has been previously described (6). pDrep1, constructed in the pGEM-3Zf(-) (Promega) backbone, was linearized at the *Hpa* I site at genomic nt 152 by overnight digestion at 37° C. The linearized plasmid was isolated using the DNA Clean and Concentrator-5 kit (Zymo Research) and used in an *in vitro* T7 RNA polymerase transcription reaction with RiboMax Large Scale RNA Production System (Promega). To produce a transcript with a totally unmethylated cap structure, G(5')ppp(5')G (New England Biolab) was used in the reaction mixture, and for a partially methylated cap, Ribo ^{7}m cap analog (Promega) was used in the reaction mixture, and for a 5' structure with no cap at all, nothing was added. In this last case the 5' terminus was 5'GAUUGUG... *In vitro* RNA transcripts were phenol/chloroform extracted and purified by chromatography through a Bio-Spin 6 column (Bio-Rad) to remove unincorporated nucleotides, and ethanol precipitated.

In vitro assay for BCoV 2'-*O*-Methyltransferase enzyme activity

For the *in vitro* assay for 2'-*O*-Methyltransferase enzyme activity, the method of Egloff et al (8) was used. Unmethylated capped, partially-methylated capped, and uncapped RNA transcripts were added to reaction mixtures at 19μM with 40mM Tris-HCl, 100μM AdoMet with 10μCi Ado[methyl-^{3}H]Met 85Ci/mmole, 5μg purified 2'-*O*-MT, 5% glycerol, and 1mM DTT. 8μl aliquots of each sample and aliquots of a control reaction containing no RNA were removed at time points 0, 10, 20, 60, 120, 180, 240, 300, and 350 minutes, mixed with 10μl of 100μg/ml BSA and 5% glycerol, and the entire 18μl was spotted onto a pre-wetted Whatman DEAE-81 paper disc. The paper discs were washed sequentially with four 1ml volumes of 200 mM ammonium bicarbonate, four 1ml volumes of water, and four 1ml volumes of ethanol, air dried, and tritium incorporation was determined by liquid scintillation counting.

Results

Comparative analysis of the BCoV 2'-*O*-Methyltransferase with other methyltransferases

By comparison with other MTs, the BCoV MT appears to be a member of the RrmJ family of MTs and shares a common structural fold with other viral MTs such as VP39 from vaccinia virus (25), a cytoplasmic DNA virus, and $NS5MT_{DV}$ from Dengue virus, a cytoplasmic positive-stranded RNA virus (8) (Table III-1). The BCoV MT is a protein of about 35 kDa and contains a predicted pocket for the binding of SAM, the methyl group donor in the methyl transfer reaction in the capping process.

Table III-1. Characteristics of viral methyltransferases. Activity has not been established for each of these enzymes. Sequence analysis predicts the activity indicated. Putative SAM binding/active site indicates the catalytic region based on sequence and amino acid comparisons with analogous sites in other MT enzymes.

Protein Name	Organism	Reaction	Methyl Donor	Putative SAM Binding/ Active Site	Comments	Ref.
Λ2	Reovirus	Guanosine-7-N MT; 2'-*O*-MT	SAM	Residues 825-888	Both MT activities observed in Reovirus cores	(26)
NS5	Dengue	2'-*O*-MT	SAM	Cleft between β strands 1 and 4	MT at N-terminal domain of RdRP	(8)
155 kDa poly-peptide	Bamboo Mosaic Virus	GTP methyl-transferase; Guanylyl-transferase	SAM	N terminal 442 amino acids	GTP moiety methylated before transfer to mRNA	(22)
VP39	Vaccinia Virus	2'-*O*-MT	SAM	Aromatic residue-lined cleft bisecting major face of protein	VP39 binds VP55	(25)
RdRp	Vesicular Stomatitis Virus	2'-*O*-MT	SAM	Motifs I, III, IV - SAM binding; IV, VI, VIII, X-catalytic site	MT activity found on RdRP	(12)
RdRP L protein	Sendai Virus	Guanosine-7-N MT; 2'-*O*-MT	SAM	C terminus of L protein	N terminus of protein has poly-merase activity	(28)

There are three general steps in the capping process (8). In the first, an RNA triphosphatase converts the 5'-triphosphate of the mRNA to a diphosphate. In the second step, a guanosine monophosphate (GMP) moiety is transferred from guanosine triphosphate (GTP) to the 5'-diphosphate RNA by a guanylyltransferase. In the third step, the reaction consists of methylating the transferred guanosine moiety by a guanine-N7-methyltransferase by the use of SAM as a methyl donor. In a second methylation reaction, the first nucleotide 3' to the triphosphate bridge is methylated by a nucleoside-2'-*O*-methyltransferase. Adjacent nucleotides located 3' to the first methylated base may also be methylated and the order of the reactions may vary. The BCoV MT is thought to function in the last of these steps in which a methyl group is transferred to the first transcribed nucleotide of the mRNA.

Various assays have been developed to measure viral methyltransferase activity. Experiments using the negative-strand RNA Sendai virus show that L protein, which has RNA-dependent RNA polymerase activity, catalyzes guanosine methylation of virus-specific mRNA. Sequence analysis of the C-terminus of this protein also suggests a possible 2'-*O*-MT domain (28). In another example, VP39 from the double-stranded DNA vaccinia virus is the sole structurally characterized viral methyltransferase that is known to have 2'-*O*-methyltransferase activity (2). We chose to use the methyltransferase assay conditions of Egloff *et al* (8) employed in a study of Dengue virus (flavivirus) MT since the BCoV MT has bioinformatics-predicted similarities in structure and function (11). The Dengue virus MT, NS5, is part of a multifunctional protein wherein the MT domain is at the N-terminus and the RdRP domain is at the C-terminus. In the Egloff assay, short potential acceptor RNA transcripts were either

capped, had a partially methylated cap (in which the transferred GMP is methylated), or were uncapped, and tritiated SAM was used as the methyl donor. In the Dengue virus MT assay, the capped and partially methylated capped transcripts were better acceptors for methylation by the 2'-*O*-Methyltransferase than the uncapped transcripts (8). This fits with the notion that the Dengue virus NS5 protein is a 2'-*O*-Methyltransferase, participating in a reaction which transfers a methyl group to the first transcribed nucleotide of the mRNA.

To produce target RNA for methylation in our assay, pDrep1 was linearized at the *Hpa* I site at nt 152 and used as a template for T7 RNA polymerase-driven *in vitro* transcription. pDrep1 contains the entire 5'- and 3'-UTRs, a portion of ORF 1a encoding protein p28, and the entire N gene fused in-frame to make a single long open reading frame. RNA transcripts were 5'-terminally capped, uncapped or contained an unmethylated cap. Tritium-labeled SAM served as the methyl donor and purified BCoV MT was the enzyme source. The abundance of tritium incorporation was the measure of MT activity.

The BCoV 2'-*O*-Methyltransferase did not methylate RNA transcripts

When tritium incorporation was measured for capped, uncapped, and unmethylated capped transcripts, no evidence of RNA methylation was found (Figure III-1). One major peak was observed for the unmethylated capped RNA at ~15 minutes, but it was similar to a peak in the negative control at ~210 minutes so the results were judged to be inconclusive.

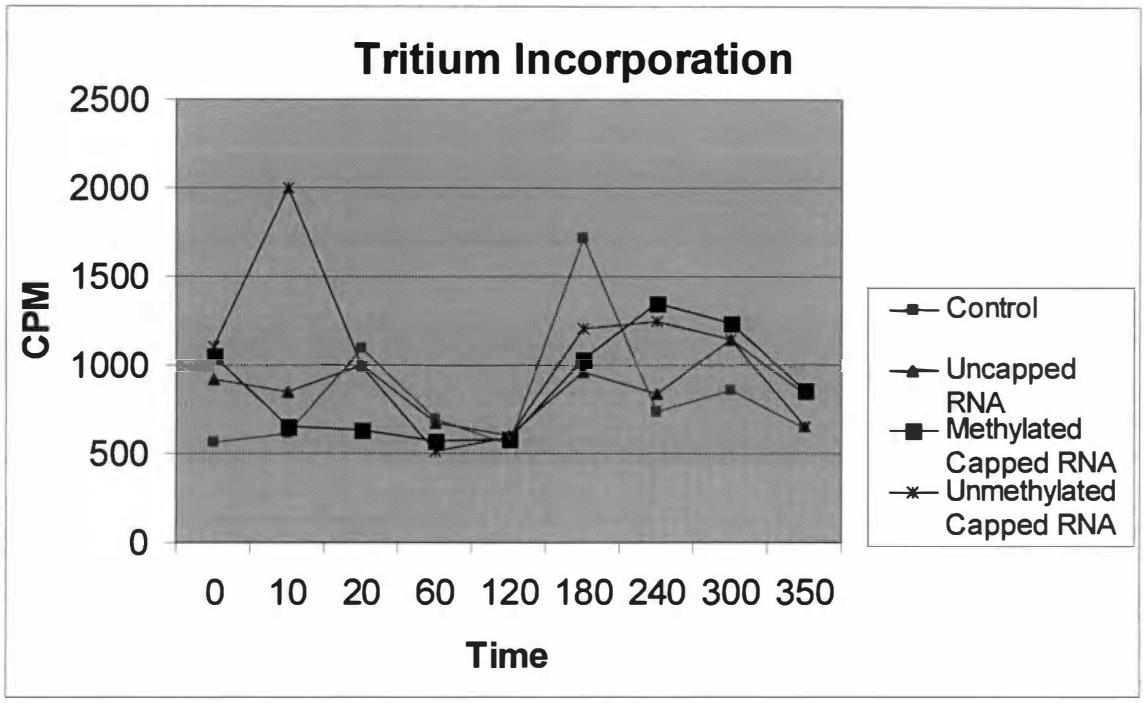

Figure III-1. Tritium incorporation of RNA transcripts. This graph shows tritium incorporation of capped, uncapped, and unmethylated capped RNA transcripts plus a negative control over a 350 minute time period. Tritium incorporation was measured by liquid scintillation counting. Five readings were taken for each sample; the highest and lowest values were discarded and the remaining 3 values were averaged to generate the graphs.

Discussion

Our results from the functional assay for BCoV 2'-*O*-MT activity indicate that the enzyme was not active under the conditions tested. One problem with the assay was the lack of a reliable level for the negative control. The negative control values should have been at background levels (around 400-500 cpm) (8) but were consistently 500 cpm and higher. This indicates that the values for the sample RNA transcripts are not reliable apparently because some of the unincorporated ^{3}H-SAM was not getting washed away. Also, one explanation for the apparent inactivity of the enzyme in this assay is that the BCoV 2'-*O*-MT may not function as an individual protein but as a polyprotein or as a membrane-bound protein complex with the other replication enzymes. Many viral proteins that function as enzymes from positive-strand RNA viruses carry out their enzymatic function as a protein precursor of the final digested product. For example, poliovirus protein CD is a precursor to the final digested products, C (a protease) and D (the RdRp). The protease functions only as a precursor CD molecule. If this is the case with the BCoV MT, using purified recombinant protein alone would not support the function of the enzyme and, therefore, would not allow for measurable methylation of RNA transcripts.

CHAPTER IV

BACTERIALLY-EXPRESSED BOVINE CORONAVIRUS 2'-*O*-METHYLTRANSFERASE BINDS *CIS*-ACTING STEM-LOOP IV IN THE 5' UNTRANSLATED REGION OF THE VIRAL GENOME

Introduction

In Chapter I are described six *cis*-acting elements for BCoV genome replication as determined through a study of the 2.2 kilobase BCoV DI RNA. These are summarized in Fig. I-3. Potentially these elements act as binding sites for the replication enzymes encoded in ORF 1b or other viral proteins involved in RNA replication. Alternatively, they could provide an assembly focus for one enzyme that then binds others cooperatively in the formation of a replication complex, which in turn leads to replication of the genome and the antigenome. For example, it has been shown that the nucleocapsid (N) protein binds stem loop III of the 5' UTR (Raman, submitted).

In this study we were interested in exploring potential interactions between the BCoV 2'-*O*-MT and *cis*-acting replication elements in the 5' UTR of the BCoV genome. For this, a series of electrophoretic mobility shift assays (EMSA) using probes representing three *cis*-acting elements in the 5' UTR were used. The absence of the MT binding to two other regions of the genome known to contain *cis*-acting elements, namely (i) the region of the genome encoding nsp 1, and (ii) the bulged stem-loop and adjacent pseudoknot in the 5'-proximal region of the 3' UTR, served as negative controls for protein binding. The assays for the MT binding to the nsp 1 region and the bulged stem-loop and adjacent pseudoknot were done collaboratively with Kortney Gustin and Agnieszka Dziduszko, respectively. It was observed that the BCoV MT binds stem-loop

IV. This is the first report of an RNA-protein interaction for a coronavirus-encoded methyltransferase.

Materials and Methods

Construction of plasmids

Construction of plasmid p225 was accomplished by using primers p225Eco (5'-GAATTCGATTGTGAGCGATTTGCGTGCGTGCATTC-3') and Hind225(+) (5'-AAGCTTGTTGATCTTCGACATTGTGACCTAT-3') in a PCR reaction with pDrep1 as a template to generate a PCR amplicon of the first 225 nucleotides of the BCoV genome. The p225Eco (5') primer was designed to have an *EcoR* I site, and the Hind225(+) (3') primer was designed to have a *Hind* III site. The PCR product was isolated from the gel with the use of Q·Biogene's GeneClean III kit. The isolated product was cloned into TOPO-XL and transformed into TOP10 electrocompetent cells (Invitrogen). Cells were plated onto plates containing 2XYT agar and kanamycin and grown overnight at 37° C. A PCR screen was done to determine the colonies positive for the plasmid, and sequencing was carried out at the UTK Molecular Biology Resource Facility to verify that the inserted sequence was wild-type. A minipreparation was made from a colony containing the wild-type plasmid using Promega's Wizard Miniprep kit. This miniprep was used in a double digestion reaction with *EcoR* I and *Hind* III. The digested insert was isolated from a 1% agarose ethidium bromide gel by GeneClean and cloned into the vector pGEM-3Zf(-) (Promega). Construction of the plasmid pSLIV has been previously described (31). pSLIV is a plasmid containing the 30-nt stem-loop IV region and flanking sequence on both the 5' and 3' ends, contained in the vector pGEM-3Zf(-).

Production of antibodies to the BCoV 2'-*O*-Methyltransferase

pGex-MT was used by the Proteintech Group, Inc. (Chicago IL) for commercial production of rabbit polyclonal antibodies to BCoV 2'-*O*-MT. We have received both preimmune and immune serum preparations as well as 1 mg of affinity column-purified fusion protein of BCoV 2'-*O*-MT from Proteintech.

***In vitro* transcription**

The plasmid p225 was linearized with *Hind* III in an overnight reaction at 37°C. The plasmid pSLIV was linearized using *Nco* I in an overnight reaction at 37° C. The linearized DNA was isolated using DNA Clean and Concentrator-5 from Zymo Research. The isolated DNA was added to a 50µl *in vitro* transcription reaction mixture to make a final concentration of 2.5µg per reaction, with 10µl 5X transcription buffer, 5µl 100mM DTT, 2.5µl acetylated BSA (2mg/ml), 1µl RNasin (40U/ul), 10µl rGTP (2.5mM), 10µl rATP (2.5mM), 10µl rCTP (2.5mM), 3µl rUTP (200µM), 12µl α ^{32}P-UTP, and 1µl T7 polymerase. The reaction mixture was incubated at 37° C for 1 hour, at which time 2.5µl RQ1 RNase-free DNase was added, and the reaction incubated for an additional 30 minutes at 37° C. 50µl sequencing stop dye was added to the reaction and the entire reaction mix was electrophoretically resolved on a denaturing urea/polyacrylamide gel at 200 volts constant voltage. The gel was exposed to film for 10 minutes and the isolated probes were cut directly from the gel and eluted overnight on a rotary tumbler at 4° C in 0.5M ammonium acetate and 1mM EDTA. The eluted probes were ethanol precipitated and resuspended in 20µl of water. The radioactive content of

the probes was measured by Cerenkov counting and the probes diluted to 1-2 X 10^4 cpm/ul.

Protein binding assays

For protein binding, essentially the conditions of Thomson (35) were used, but with the addition of yeast tRNA (16) and heparin (3). To 20 µl of a mixture of 5mM HEPES, 40mM KCl, 2mM $MgCl_2$, 4% glycerol, 2mM DTT, yeast tRNA (1µg/µl), heparin (0.25µg/µl), PMSF (1X), and RNasin inhibitor (20U/µl) was added 0.35 to 5µg of purified MT protein and the reaction was incubated at 37° C for 10 minutes. The yeast tRNA was added as a non-specific competitive inhibitor. To certain reactions, 5 to 15µg of cold competitor RNA was added, and the reactions were incubated for an additional 10 minutes at 37°C. To this mixture was added 1µl of radioactive probe and the reaction mixture was incubated at 37° C for 10 minutes.

Electrophoretic mobility shift assays

For electrophoretic separation of RNA-protein complexes, 2 µl of 50% glycerol was added to the probe-protein mixture and electrophoresis was carried out on a native gel of 5% polyacrylamide-5% glycerol for the stem-loop IV construct and 4% polyacrylamide-5% glycerol for the p225 construct at 4° C with 0.5×TBE running buffer (1×TBE= 90 mM Tris HCl, 90 mM boric acid, 2 mM EDTA) for approximately 4 h at 100 volts constant voltage.

Results

The Bovine Coronavirus 2'-*O*-Methyltransferase binds stem-loop IV of the 5' untranslated region of the genome

To investigate the binding properties of the MT, a series of gel shift assays were performed using the purified MT and radiolabeled probes encompassing the entire 5' UTR (p225) or only stem-loop IV (pSLIV) (Figure IV-1). In the first assay, the binding of the MT to a 225-nt long ^{32}P-UTP labeled probe, a probe encompassing the first 225 nts of the BCoV genome which included the 210-nt 5' UTR and 15 adjacent nts, was done, and it was observed that this interaction produced a shift, even in the presence of yeast tRNA, a non-specific competitive inhibitor (Figure IV-2, *A*).

To begin to address the question of which specific elements, if any, within the 5' UTR the MT might be binding, the probe encompassing only stem-loop IV and synthesized from pSLIV (Fig. IV-1 *B*) was used. This construct contains the 30-nt stem-loop IV flanked by 36 nts upstream and 47 nts downstream. For this experiment, the pSLIV plasmid was linearized at the *Nco* I site at genomic nt 262, producing a 113-nt transcript upon *in vitro* transcription. The radiolabeled probe was incubated with MT and the reaction was resolved by polyacrylamide gel electrophoresis on a non-denaturing gel. The results shown in Fig. IV-2, *D* and *E* demonstrate that the MT binds stem-loop IV with enough affinity to cause a retardation of electrophoretic migration, a gel shift.

A probe representing the nsp 1 coding region, made by Kortney Gustin (probe P1 in Fig. IV-2, *B*), and a probe representing the bulged stem-loop and adjacent pseudoknot in the 3' UTR, made by Agnieszka Dziduszko (probe PK in Fig. IV-2, *C*), did not bind the MT. This is illustrated by the absence of a gel shift with the MT in Figure IV-2, *B*

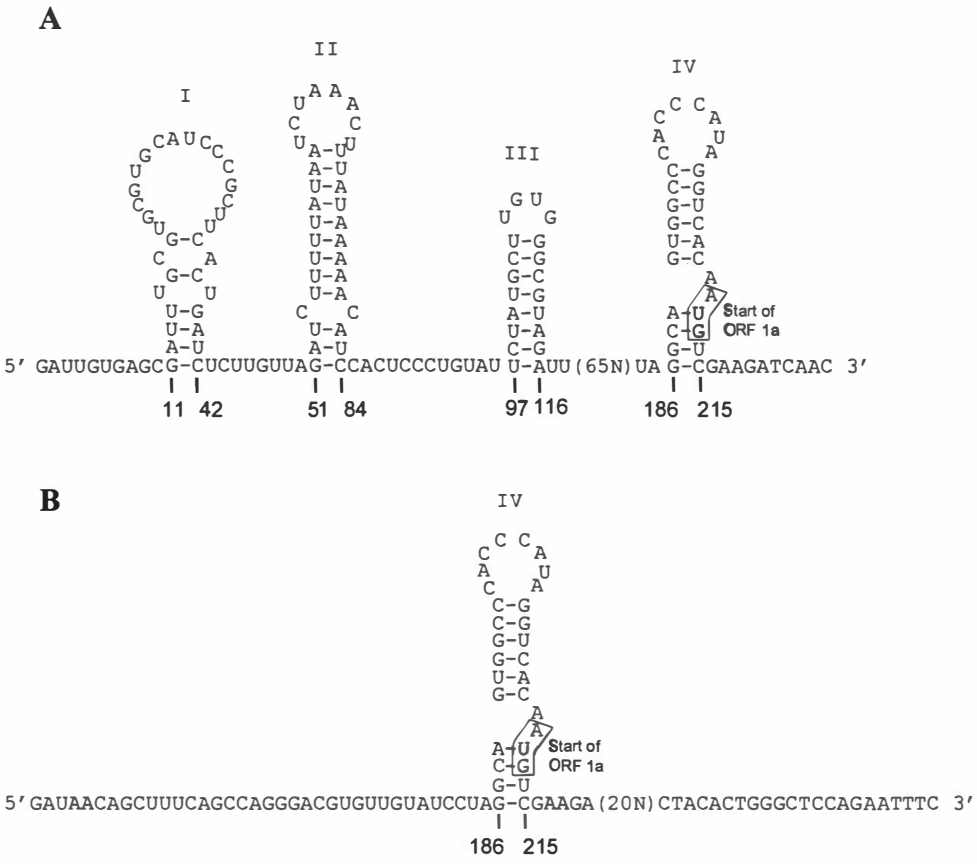

Figure IV-1. Structure of the probes used in the EMSAs. *A*. p225 probe made by T7 RNA polymerase after linearizing plasmid p225 at the *Hind* III site. *B*. pSLIV probe made by T7 RNA polymerase after linearizing plasmid pSLIV at the *Nco* I site. Numbers at the bottom of the figure indicate genomic nucleotide position.

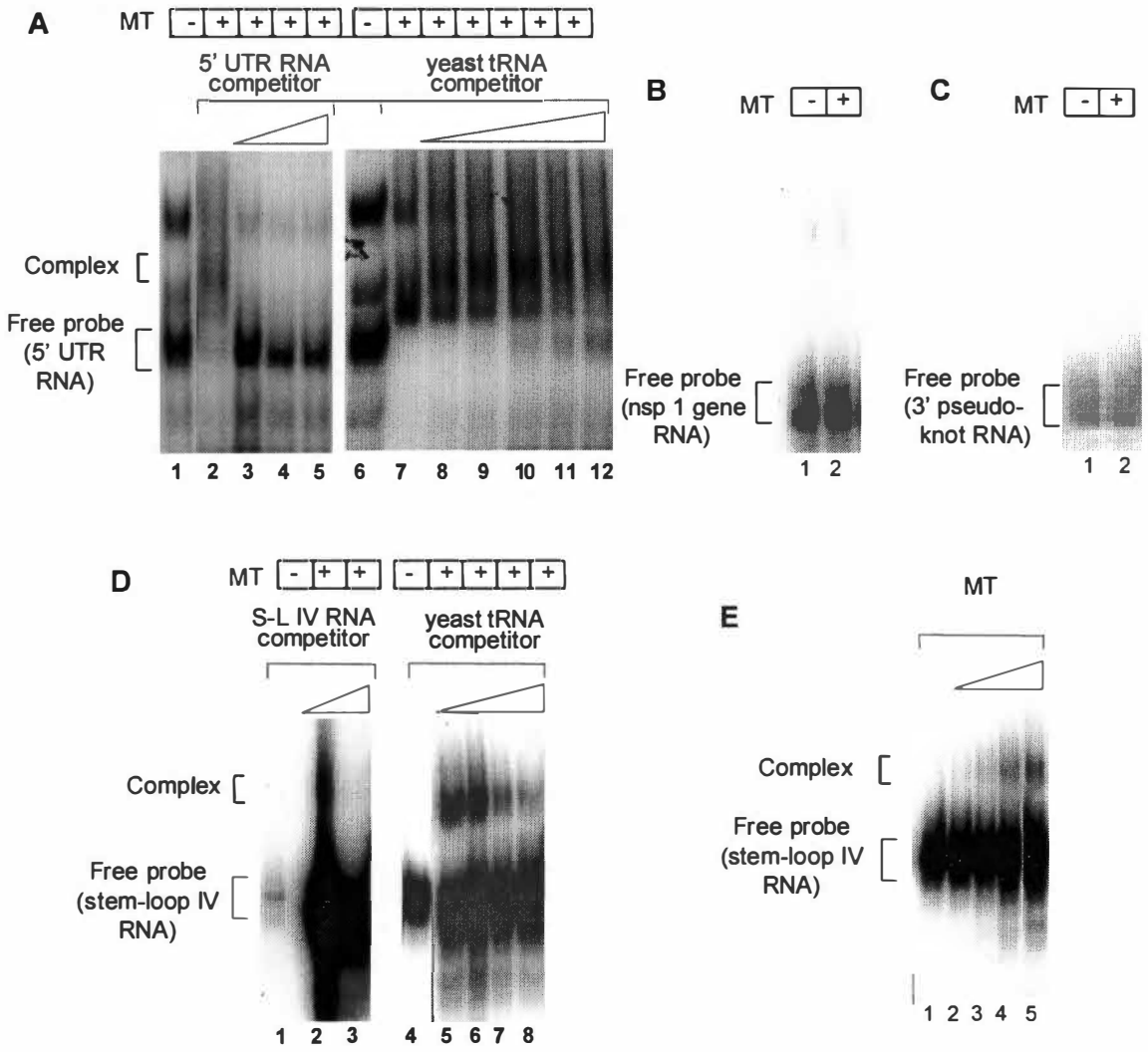

Figure IV-2. Gel shift assays. *A.* Assay with probe p225. Lanes 1 and 6, free probe; lane 2, p225 with 5μg MT; lanes 3, 4, and 5, with 5, 10, and 15μg of cold competitor RNA, respectively; lanes 7, 8, 9, 10, 11, 12, with 0, 5, 10, 15, 20, and 25μg of yeast tRNA, respectively. *B.* Assay with probe P1, containing stem-loops V and VI. Lane 1, free probe; lane 2, with 1:1000 ratio of probe to protein. *C.* Assay with probe PK, containing the 5' proximal pseudoknot of the 3' UTR. Lane 1, free probe; lane 2, with 1:100 ratio of probe to protein. *D.* Competition assay with pSLIV probe. Lanes 1 and 4, free probe; lanes 2 and 3, with 5 and 10μg of cold competitor RNA, respectively. Lanes 5, 6, 7, and 8, with 2.5, 5, 10, and 20μg tRNA, respectively. *E.* Assay with pSLIV probe. Lane 1, free probe; Lanes 2, 3, 4, 5, with 0.35, 1.75, 3.5, and 5μg of MT, respectively. tRNA was used as a non-specific competitive inhibitor. Unless otherwise indicated, all lanes contain 5μg MT and 20μg tRNA.

and *C*. P1 is a 134-nt probe containing stem-loops V and VI, and PK is a 147-nt probe containing the 5'-proximal bulged stem-loop and adjacent pseudoknot in the 3' UTR.

Discussion

The results of the electrophoretic mobility shift assays performed here indicate that the BCoV 2'-*O*-Methyltransferase binds within the 5' UTR of the genome, and that at least one of the elements responsible for the binding within this region is stem-loop IV. The binding of the MT to an element in the 5' UTR of the genome is not unexpected since it is predicted to function in the capping of the genome and subgenomic mRNAs, and these occur at the genome 5' terminus. Many questions need yet to be answered about the BCoV MT concerning its function. What does it do? Does it, in fact, act to methylate the cap structure, or might it methylate at other sites? For example, does it methylate bases internally within the genome that might serve as signals for other processes? It is still not clear what signals the RdRp to template switch at internal sites in order to place a leader onto sg mRNAs (20, 29). Conceivably, a methyl group might signal such a template switch. Does the MT per chance methylate a viral protein and hence regulate its function?

In a separate set of experiments performed by Agnieszka Dziduszko with the MT described in this study, it was found that the MT binds a region near the 3' end of the genome. Preliminary data suggest it binds at or near the octameric sequence, GGAAGAGC, found universally in coroanvirus genomes at approximately 70 nts in from the 3' end of the genome. Mutations in the octameric sequence destroy the replicating

ability of the BCoV DI RNA, pDrep1 (Hung-Yi Wu, unpublished), suggesting it has importance as a replication signal. Perhaps the binding of the MT to the 3' end of the genome is involved in genome cyclization, a postulated requirement for replication of many positive-stranded RNA viruses (1, 14, 15, 19, 38).

Ultimately it will be important to determine where, within the BCoV genome, the MT does bind and, equally important, with what other viral proteins the MT interacts. It will also be important to determine for both its enzymatic and binding functions whether the MT acts as a final cleaved product, as synthesized here, or whether it acts as an uncleaved precursor, of which there are at least 4 possibilities: MT-EndoN, MT-EndoN-ExoN, MT-EndoN-ExoN-Hel, and MT-EndoN-ExoN-Hel-RdRP.

CHAPTER V

FUTURE DIRECTIONS

Several questions raised by the results of this study and made approachable by the reagents produced are the following. (i) The other *cis*-acting elements within the 5' UTR (the sequence represented by stem-loops I and II and stem-loop III) need to be explored as possible binding sites for the MT. A stem-loop III construct will need to be constructed and used. (ii) It will be important to determine whether the binding of MT to stem-loop IV is dependent upon its higher-order structure, as is its *cis*-function for replication. (iii) It will be important to determine which, if any, of the 5'-proximal *cis*-acting replication elements are bound by the other enzymes encoded in ORF 1b. (iv) It will be important to determine which viral proteins, if any, interact with the bound proteins in a secondary manner. How are they assembled to make the replication complexes? Gel shift assays using all the replication enzymes in varying combinations can be performed in order to determine the sequence of replication complex assembly. Additionally, the antibodies produced against these proteins can be used in pull-down assays to study assembly. (v) It will be important to determine where the MT might bind in other regions of the genome. My laboratory colleague Agnieszka Dziduszko has made the observation that the MT binds the octamer-containing bulged stem-loop in the 3' UTR. Is this playing a role in replication complex assembly or in genome cycilization? (vi) It will be important to determine the details of the MT-RNA interactions with such studies that might entail crystal structure analysis of the BCoV 2'-*O*-Methyltransferase. This would be extremely helpful in discovering the enzyme's mechanism of action, its

binding site at high resolution, and ultimately in designing chemotherapeutic drugs targeted to the MT. It's thought that the MT binds to the 5' UTR in order to aid in the 5' capping reaction; however, mutational analyses will need to be done to determine specificity. In addition, the other *cis*-acting elements will need to be tested in order to determine the details of the MT binding to the 5' UTR and catalyzing the capping reaction. The model presented in Figure V-1 is a possible scenario for the binding of the MT to stem-loop IV. However, more experiments (described above) will need to be performed in order to more fully elucidate the purpose of the MT-RNA interactions. Attempts to crystallize the MT in collaboration with the Christopher Dealwis laboratory at the University of Tennessee in Knoxville, are currently underway. (vii) Efforts to clone wild-type RdRP and Hel and produce antibodies to them should continue such that a systematic characterization of replication complex structure and function can be made.

Figure V-1. Model representing binding of the BCoV MT to *cis*-acting stem-loop IV. Stem-loop IV may be providing a binding target in order for the MT to perform the methyltransfer reaction. *In vivo*, the RNA structure of the mRNA may be folding into a tertiary configuration in which the stem-loop IV bound MT is in close proximity to the 5' terminus of the transcript.

REFERENCES

REFERENCES

1. **Ackermann, M., and R. Padmanabhan.** 2001. De novo synthesis of RNA by the dengue virus RNA-dependent RNA polymerase exhibits temperature dependence at the initiation but not elongation phase. J Biol Chem **276:**39926-37.
2. **Barbosa, E., and B. Moss.** 1978. mRNA(nucleoside-2'-)-methyltransferase from vaccinia virus. Characteristics and substrate specificity. J Biol Chem **253:**7698-702.
3. **Blyn, L. B., J. S. Towner, B. L. Semler, and E. Ehrenfeld.** 1997. Requirement of poly(rC) binding protein 2 for translation of poliovirus RNA. J Virol **71:**6243-6.
4. **Brian, D. A., and Baric, R.S.** 2005. Coronavirus genome structure and replication. Current Topics in Microbiology and Immunology **287:**1-30.
5. **Chang, R. Y., and D. A. Brian.** 1996. cis Requirement for N-specific protein sequence in bovine coronavirus defective interfering RNA replication. J Virol **70:**2201-7.
6. **Chang, R. Y., M. A. Hofmann, P. B. Sethna, and D. A. Brian.** 1994. A cis-acting function for the coronavirus leader in defective interfering RNA replication. J Virol **68:**8223-31.
7. **Chouljenko, V. N., X. Q. Lin, J. Storz, K. G. Kousoulas, and A. E. Gorbalenya.** 2001. Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. J Gen Virol **82:**2927-33.
8. **Egloff, M. P., D. Benarroch, B. Selisko, J. L. Romette, and B. Canard.** 2002. An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. Embo J **21:**2757-68.
9. **Enjuanes, L., D. Brian, D. Cavanagh, K. Holmes, M. M. C. Lai, H. Laude, P. Masters, P. Rottier, S. G. Siddell, W. J. M. Spaan, F. Taguchi, and P. Talbot.** 2000. Coronaviridae. *In* C. M. F. M. H. V. van Regenmortel, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, D. J. McGeoch, J. Maniloff, M. A. Mayo, C. R. Pringle, and R. B. Wickner. (ed.), Virus taxonomy: classification and nomenclature of viruses. Academic Press, San Diego.
10. **Enjuanes, L., W. Spaan, E. Snijder, and D. Cavanagh.** 2000. Nidovirales. *In* C. M. F. M. H. V. van Regenmortel, D. H. L. Bishop, E. B. Carsten, M. K. Estes, S. M. Lemon, D. J. McGeoch, J. Maniloff, M. A. Mayo, C. R. Pringle, and R. B. Wickner (ed.), Virus taxonomy: classification and nomenclature of viruses. Academic Press, San Diego.
11. **Feder, M., J. Pas, L. S. Wyrwicz, and J. M. Bujnicki.** 2003. Molecular phylogenetics of the RrmJ/fibrillarin superfamily of ribose 2'-O-methyltransferases. Gene **302:**129-38.
12. **Ferron, F., S. Longhi, B. Henrissat, and B. Canard.** 2002. Viral RNA-polymerases -- a predicted 2'-O-ribose methyltransferase domain shared by all Mononegavirales. Trends Biochem Sci **27:**222-4.

13. **Foley, J. E., and C. Leutenegger.** 2001. A review of coronavirus infection in the central nervous system of cats and mice. J Vet Intern Med **15:**438-44.
14. **Frolov, I., R. Hardy, and C. M. Rice.** 2001. Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. Rna **7:**1638-51.
15. **Gallie, D. R.** 2002. Protein-protein interactions required during translation. Plant Mol Biol **50:**949-70.
16. **Gamarnik, A. V., and R. Andino.** 1997. Two functional complexes formed by KH domain containing proteins with the 5' noncoding region of poliovirus RNA. Rna **3:**882-92.
17. **Goebel, S. J., B. Hsue, T. F. Dombrowski, and P. S. Masters.** 2004. Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. J Virol **78:**669-82.
18. **Herndon, R. M., D. E. Griffin, U. McCormick, and L. P. Weiner.** 1975. Mouse hepatitis virus-induced recurrent demyelination. A preliminary report. Arch Neurol **32:**32-5.
19. **Herold, J., and R. Andino.** 2001. Poliovirus RNA replication requires genome circularization through a protein-protein bridge. Mol Cell **7:**581-91.
20. **Hofmann, M. A., P. B. Sethna, and D. A. Brian.** 1990. Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture. J Virol **64:**4108-14.
21. **Hogue, B. G., B. King, and D. A. Brian.** 1984. Antigenic relationships among proteins of bovine coronavirus, human respiratory coronavirus OC43, and mouse hepatitis coronavirus A59. J Virol **51:**384-8.
22. **Huang, Y. L., Y. T. Han, Y. T. Chang, Y. H. Hsu, and M. Meng.** 2004. Critical residues for GTP methylation and formation of the covalent m7GMP-enzyme intermediate in the capping enzyme domain of bamboo mosaic virus. J Virol **78:**1271-80.
23. **Ivanov, K. A., V. Thiel, J. C. Dobbe, Y. van der Meer, E. J. Snijder, and J. Ziebuhr.** 2004. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J Virol **78:**5619-32.
24. **Kienzle, T. E., S. Abraham, B. G. Hogue, and D. A. Brian.** 1990. Structure and orientation of expressed bovine coronavirus hemagglutinin-esterase protein. J Virol **64:**1834-8.
25. **Lockless, S. W., H. T. Cheng, A. E. Hodel, F. A. Quiocho, and P. D. Gershon.** 1998. Recognition of capped RNA substrates by VP39, the vaccinia virus-encoded mRNA cap-specific 2'-O-methyltransferase. Biochemistry **37:**8564-74.
26. **Luongo, C. L., C. M. Contreras, D. L. Farsetta, and M. L. Nibert.** 1998. Binding site for S-adenosyl-L-methionine in a central region of mammalian reovirus lambda2 protein. Evidence for activities in mRNA cap methylation. J Biol Chem **273:**23773-80.
27. **Nelson, G. W., S. A. Stohlman, and S. M. Tahara.** 2000. High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. J Gen Virol **81:**181-8.

28. **Ogino, T., M. Kobayashi, M. Iwama, and K. Mizumoto.** 2005. Sendai virus RNA-dependent RNA polymerase L protein catalyzes cap methylation of virus-specific mRNA. J Biol Chem **280:**4429-35.
29. **Ozdarendeli, A., S. Ku, S. Rochat, G. D. Williams, S. D. Senanayake, and D. A. Brian.** 2001. Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. J Virol **75:**7362-74.
30. **Raman, S., P. Bouma, G. D. Williams, and D. A. Brian.** 2003. Stem-loop III in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. J Virol **77:**6720-30.
31. **Raman, S., and D. A. Brian.** 2005. Stem-loop IV in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. J Virol **79:**12434-46.
32. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. Spaan, and A. E. Gorbalenya.** 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol **331:**991-1004.
33. **Spagnolo, J. F., and B. G. Hogue.** 2000. Host protein interactions with the 3' end of bovine coronavirus RNA and the requirement of the poly(A) tail for coronavirus defective genome replication. J Virol **74:**5053-65.
34. **Tanaka, R., Y. Iwasaki, and H. Koprowski.** 1976. Intracisternal virus-like particles in brain of a multiple sclerosis patient. J Neurol Sci **28:**121-6.
35. **Thomson, A. M., J. T. Rogers, C. E. Walker, J. M. Staton, and P. J. Leedman.** 1999. Optimized RNA gel-shift and UV cross-linking assays for characterization of cytoplasmic RNA-protein interactions. Biotechniques **27:**1032-9, 1042.
36. **Williams, G. D., R. Y. Chang, and D. A. Brian.** 1999. A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication. J Virol **73:**8349-55.
37. **Wu, H. Y., J. S. Guy, D. Yoo, R. Vlasak, E. Urbach, and D. A. Brian.** 2003. Common RNA replication signals exist among group 2 coronaviruses: evidence for in vivo recombination between animal and human coronavirus molecules. Virology **315:**174-83.
38. **You, S., and R. Padmanabhan.** 1999. A novel in vitro replication system for Dengue virus. Initiation of RNA synthesis at the 3'-end of exogenous viral RNA templates requires 5'- and 3'-terminal complementary sequence motifs of the viral RNA. J Biol Chem **274:**33714-22.
39. **Ziebuhr, J.** 2005. The coronavirus replicase. Curr Top Microbiol Immunol **287:**57-94.
40. **Zuker, M.** 2000. Calculating nucleic acid secondary structure. Curr Opin Struct Biol **10:**303-10.

VITA

Tara Beth Tucker was born in Manchester, TN in 1979. After graduation from Coffee County Central High School in 1997, she attended Vanderbilt University in Nashville, TN. In May 2001, she received her Bachelor of Science degree in Biology. She entered the Master's program in Microbiology at the University of Tennessee in Fall 2002. She performed coronavirus research in the laboratory of Dr. David Brian, and received her Master's degree in Fall 2005.